

On Bernstein Type Inequalities for Stochastic Integrals of Multivariate Point Processes

Hanchao Wang¹

Zhongtai Security Institute for Financial Studies, Shandong University, Jinan, 250100, PRC

Zhengyan Lin, Zhonggen Su

School of Mathematical Sciences, Zhejiang University, Hangzhou, 310027, PRC

Abstract

We consider the stochastic integrals of multivariate point processes and study their concentration phenomena. In particular, we obtain a Bernstein type of concentration inequality through Doléans-Dade exponential formula and a uniform exponential inequality using a generic chaining argument. As applications, we obtain an upper bound for a sequence of discrete time martingales indexed by a class of functionals, and so derive the rate of convergence for nonparametric maximum likelihood estimators, which is an improvement of earlier work of van de Geer.

Keywords: Bernstein inequality, Doléans-Dade exponential formula, Generic chaining method, Multivariate point process.

1. Introduction

There have been a lot of research activities around phenomena of measure concentration in the past decades. The reader is referred to excellent books like Ledoux and Talagrand [12], Ledoux [11] and nice paper like Talagrand

¹Corresponding author, email: wanghanchao@sdu.edu.cn.

[14] for remarkable results and powerful methods. A primary purpose of the present paper is to establish a Bernstein type exponential concentration inequality for stochastic integrals of multivariate point processes.

For sake of statement, we will begin with a classical Bernstein inequality for sums of independent random variables. Assume that $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space so large that we can construct all random objects of interest in it. Let ξ_1, ξ_2, \dots be a sequence of centered independent random variables with finite variance, and denote $S_n = \xi_1 + \xi_2 + \dots + \xi_n$. If there exists a certain constant $a > 0$ such that

$$\mathbf{E}[|\xi_k|^p] \leq \frac{p!a^{p-2}}{2} \mathbf{E}[\xi_k^2], \quad p \geq 2 \quad (1.1)$$

then

$$\mathbf{P}(|S_n| \geq nx) \leq 2 \exp\left(-\frac{nx^2}{2(K + ax)}\right) \quad (1.2)$$

for all $x > 0$ and for all K satisfying $\text{Var}(S_n) = \sum_{i=1}^n \mathbf{E}\xi_i^2 \leq K$.

(1.1) was due to Bernstein [6], and so (1.1) is now referred as Bernstein condition. Since then various extensions and improvement have appeared in literature, among which are Bennett [1, 2], Hoeffding [9], Freedman [8], Bentkus [4, 5], Fan et al. [7]. A very recent nice book is Bercu et al [3] which gives a very clear exposition on concentration inequalities for sums of independent random variables and martingales.

An important extension of Bernstein inequality is to both discrete time martingales and continuous time martingales. In particular, Freedman [8] first obtained the Bernstein inequality of discrete time martingales with bounded jumps, and then Shorack and Wellner [13] extended Freedman's result to continuous time martingales. More precisely, let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbf{P})$ be a stochastic basis, $\{M_t\}_{t \geq 0}$ be a locally square integrable martingale with respect to the filtration $\{\mathcal{F}_t\}_{t \in [0, T]}$ with $M_0 = 0$. Denote the jump by $\Delta M_t = M_t - M_{t-}$ and the predictable variation by $V_t = \langle M, M \rangle_t$, $t > 0$. Assume that

$$|\Delta M_t| \leq K, \quad t > 0 \quad (1.3)$$

for a positive constant K . Then for each $x, y > 0$,

$$\mathbb{P}(M_t \geq x \text{ and } V_t \leq y^2 \text{ for some } t) \leq \exp\left(-\frac{x^2}{2(xK + y^2)}\right). \quad (1.4)$$

The bounded jump assumption (1.3) can be relaxed. In fact, van de Geer [17] improved the above result under Bernstein condition. For each $m \geq 2$, consider the process $\{\sum_{s \leq t} |\Delta M_s|^m\}_{t \geq 0}$ and its predictable compensator $\{V_{m,t}\}_{t \geq 0}$. If there exist a constant $K > 0$ and a predictable process $\{R_t\}_{t \geq 0}$ such that

$$V_{m,t} \leq \frac{m!}{2} K^{m-2} R_t, \quad t \geq 0, \quad (1.5)$$

then for each $x, y > 0$,

$$\mathbb{P}(M_t \geq x \text{ and } V_t \leq y^2 \text{ for some } t) \leq \exp\left(-\frac{x^2}{2(xK + y^2)}\right). \quad (1.6)$$

We remark that any locally square integrable martingale can be represented as the sum of continuous local martingale and pure jump local martingale. The nonzero continuous local martingale part indeed played a crucial role in the proof of both (1.4) and (1.6). Now it is natural to ask what happens for a pure jump local martingale. It is an interesting and challenging mathematical problem to establish a concentration inequality for general pure jump local martingales. We shall restrict ourselves to stochastic integrals of multivariate point processes.

Let (E, \mathcal{E}) be a Blackwell space. Assume that $\{T_k\}_{k \geq 1}$ be a sequence of strictly increasing positive random variables, $\{X_k\}_{k \geq 1}$ a sequence of E -valued random variables and X_k is measurable with respect to \mathcal{F}_{T_k} for each $k \geq 1$. A multivariate point process is an integer-valued random variables defined by

$$\mu(dt, dx) = \sum_{k \geq 1} 1_{\{T_k < \infty\}} \varepsilon_{(T_k, X_k)}(dt \times dx). \quad (1.7)$$

We note that Poisson point process and compound Poisson point process are classic and well-studied examples of multivariate point processes. We

shall be interested in stochastic integrals of a predictable process with respect to the measure μ . Let ν be the predictable compensator of μ and assume that ν admits the disintegration

$$\nu(\omega, dt, dx) = dA_t(\omega)K_{\omega,t}(dx). \quad (1.8)$$

where K is a transition probability from $(\Omega \times [0, T], \mathcal{P})$ into (E, \mathcal{E}) , A is an increasing càdlàg predictable process. Denote $a_t = \nu(\{t\} \times \mathbb{R})$. It is easy to see that the process $a \equiv 0$ if μ is a Lévy point process. However, what we are more interested in the case $a \neq 0$, namely $\Delta A_t \neq 0$.

Given a predictable function W on $\tilde{\Omega}$, $\tilde{\Omega} = \Omega \times \mathbb{R}_+ \times \mathbb{R}$, define the stochastic integral

$$\hat{W}_t = \int_{\mathbb{R}} W(t, x) \nu(\{t\} \times dx). \quad (1.9)$$

In addition, put

$$\Xi(W)_t = \max\{0, (W - \hat{W})\} * \nu_t + (1 - a_s) \max\{0, -\hat{W}_t\}. \quad (1.10)$$

and

$$C(W)_t = \langle W * (\mu - \nu), W * (\mu - \nu) \rangle_t. \quad (1.11)$$

An easy computation, see Chapter 2 of Jacod and Shiryaev [10], implies

$$C(W)_t = (W - \hat{W})^2 * \nu_t + \sum_{s \leq t} (1 - a_s) (\hat{W}_s)^2. \quad (1.12)$$

Motivated by (1.12), we introduce the following quantities

$$Q(W, m)_t = \max\{0, (W - \hat{W})\}^m * \nu_t + \sum_{s \leq t} (1 - a_s) \max\{0, -\hat{W}_s\}^m, \quad m \geq 3. \quad (1.13)$$

The Bernstein inequality for $W * (\mu - \nu)$ reads as follows

Theorem 1.1. *Suppose that for all $t > 0$ and some $0 < K < \infty$*

$$\Xi(W)_t \leq K, \quad Q(W, m)_t \leq \frac{m!}{2} K^{m-2} C(W)_t, \quad m \geq 3. \quad (1.14)$$

Then for each $x > 0$, $y > 0$,

$$P(|W * (\mu - \nu)_t| \geq x \text{ and } C(W)_t \leq y^2 \text{ for some } t) \leq \exp\left(-\frac{x^2}{2(xK + y^2)}\right). \quad (1.15)$$

The proof of Theorem 1.1 will be given in Section 2. A key ingredient is Doléans-Dade exponential formula for semimartingales with given predictable characteristics.

Next let us turn to consider the uniform bound for a family of stochastic integrals of predictable processes with respect to multivariate point process. Let (Ψ, d) be a metric space, $\mathcal{W} = \{W^\psi : \psi \in \Psi\}$ a family of predictable functions on $\Omega \times \mathbb{R}_+ \times \mathbb{R}$.

Fix a $T > 0$. We denote $X^\psi = W^\psi * (\mu - \nu)_T$ and define two metrics as follows

$$d_1(\psi_1, \psi_2) = \|\Xi(W^{\psi_1} - W^{\psi_2})_T\|_\infty, \quad (1.16)$$

$$d_2(\psi_1, \psi_2) = \sqrt{\|C(W^{\psi_1} - W^{\psi_2})_T\|_\infty} \quad (1.17)$$

where $\|\cdot\|_\infty$ stands for norm of L^∞ .

By Theorem 1.1, one easily can obtain

$$P(|X^{\psi_1} - X^{\psi_2}| > x) \leq \exp\left(-\frac{x^2}{2(d_2^2(\psi_1, \psi_2) + d_1(\psi_1, \psi_2)x)}\right). \quad (1.18)$$

As known to us, (1.18) is a certain increment condition. We can further derive a uniform inequality for $\sup_{\psi \in \Psi} X^\psi$ using a generic chaining method as in Talagrand [15]. To this end, we need to introduce more notations. For a given metric space (Ψ, d) , an increasing sequence $(\mathcal{A}_n)_{n \geq 1}$ of partitions of Ψ is called as admissible sequence if $\#\mathcal{A}_n \leq 2^{2^n}$. Denote by $A_n(\psi)$ the unique element of (\mathcal{A}_n) containing ψ , and denote by $\Upsilon_d(A_n(\psi))$ the diameter of $A_n(\psi)$ under d . In addition, let

$$\gamma_\alpha(\Psi, d) = \inf \sup_{\psi \in \Psi} \sum_{n \geq 0} 2^{n/\alpha} \Upsilon_d(A_n(\psi)), \quad (1.19)$$

where the infimum is taken over all admissible sequences. We can now state a uniform inequality for $\sup_{\psi \in \Psi} X^\psi$ in terms of γ_1 and γ_2 .

Theorem 1.2. *Suppose that for all $\psi \in \Psi$ and some $0 < K < \infty$*

$$\Xi(W^\psi)_T \leq K, \quad C(W^\psi)_T^m \leq \frac{m!}{2} K^{m-2} C(W^\psi)_T, \quad m \geq 3. \quad (1.20)$$

Then we have

$$P\left(\sup_{\psi \in \Psi} |X^\psi| \geq Cu(\gamma_2(\Psi, d_2) + \gamma_1(\Psi, d_1))\right) \leq C \exp\left(-\frac{u}{2}\right). \quad (1.21)$$

Moreover, it follows

$$E \sup_{\psi \in \Psi} X^\psi \leq C(\gamma_2(\Psi, d_2) + \gamma_1(\Psi, d_1)). \quad (1.22)$$

The proof of Theorem 1.2 will also be given in Section 2. As applications, we will obtain a Bernstein type exponential inequality for a class of functional index empirical processes and so derive a convergence rate for nonparametric maximum likelihood estimators. This is the content of Section 3.

2. Proofs of Theorems 1 and 2

Proof of Theorem 1.1 Clearly, it follows

$$P(|W * (\mu - \nu)_t| \geq x) \leq P(|W| * (\mu - \nu)_t \geq x).$$

So without loss of generality, we can and do assume $W > 0$. For simplicity of notation, put

$$X_t = W * (\mu - \nu)_t,$$

and

$$S(\lambda)_t = \int_0^t \int_{\mathbb{R}} (e^{\lambda(W - \hat{W})} - 1 - \lambda(W - \hat{W})) \nu(ds, dx) + \sum_{s \leq t} (1 - a_s)(e^{-\lambda \hat{W}_s} - 1 + \lambda \hat{W}_s),$$

where $0 < \lambda < \frac{1}{K}$. Note for any semimartingale Y , the Doléans-Dade exponential is

$$\mathcal{E}(Y)_t = e^{Y_t - Y_0 - \frac{1}{2} \langle Y^c, Y^c \rangle_t} \prod_{s \leq t} (1 + \Delta Y_s) e^{-\Delta Y_s}.$$

Since

$$\Delta S(\lambda)_t = \frac{\int e^{\lambda W} \nu(\{t\}, dx) - a_t e^{\lambda \hat{W}_t} + (1 - a_t)(1 - e^{\lambda \hat{W}_t})}{e^{\lambda \hat{W}_t}}$$

and

$$\int e^{\lambda W} \nu(\{t\}, dx) + 1 - a_t > 0,$$

we can obtain

$$\Delta S(\lambda)_t > -1$$

for all $t > 0$.

We shall first show the the process $\left(e^{\lambda X}/\mathcal{E}(S(\lambda))\right)_{t \geq 0}$ is a local martingale. For X , the jump part of X is

$$\Delta X = (W - \hat{W})1_D - \hat{W}1_{D^c}, \quad (2.1)$$

where D is the thin set, which is exhausted by $\{T_n\}_{n \geq 1}$.

We denote by μ^X the jump measure of X . Let ν^X be the predictable compensator of μ^X , and

$$x * \nu_t^X = (W - \hat{W}) * \nu_t - \sum_{s \leq t} \hat{W}_s(1 - a_s). \quad (2.2)$$

The Itô formula yields

$$\begin{aligned} e^{\lambda X_t} &= 1 + \lambda e^{\lambda X_{t-}} \cdot X + \sum_{s \leq t} (e^{\lambda X_s} - e^{\lambda X_{s-}} - \lambda e^{\lambda X_{s-}} \Delta X_s) \\ &= 1 + \lambda e^{\lambda X_{t-}} \cdot X + e^{\lambda X_{t-}} \cdot (e^{\lambda x} - 1 - \lambda x) * \mu^X. \end{aligned}$$

Furthermore,

$$S(\lambda) = (e^{\lambda x} - 1 - \lambda x) * \nu^X. \quad (2.3)$$

We obtain that

$$e^{\lambda X} - e^{\lambda X_{t-}} \cdot S(\lambda) \quad (2.4)$$

is a local martingale. Set $H = e^{\lambda X}$, $G = \mathcal{E}(S(\lambda))$, $A = S(\lambda)$ and $f(h, g) = \frac{h}{g}$. The Itô formula yields

$$\begin{aligned} f(H, G) &= 1 + \frac{1}{G_-} \cdot H - \frac{H_-}{G_-^2} \cdot G \\ &\quad + \sum_{s \leq \cdot} \left(\Delta f(H, G)_s - \frac{\Delta H_s}{G_{s-}} + \frac{f(H, G)_{s-}}{G_{s-}} \Delta G_s \right). \end{aligned}$$

Let $N_1 = H - H_- \cdot A$, and note N_1 is also a local martingale. We have

$$\frac{1}{G_-} \cdot H = \frac{1}{G_-} \cdot N_1 + \frac{H_-}{G_-} \cdot S(\lambda). \quad (2.5)$$

By the definition of $\mathcal{E}(A)$, we have $G = 1 + G_- \cdot A$, thus

$$\frac{H_-}{G_-^2} \cdot G = \frac{H_-}{G_-} \cdot A. \quad (2.6)$$

Then

$$\frac{1}{G_-} \cdot H - \frac{H_-}{G_-^2} \cdot G = \frac{1}{G_-} \cdot N_1. \quad (2.7)$$

Noting that $\Delta G = G_- \Delta A$, $\Delta N_1 = \Delta H - H_- \Delta A$, we have

$$\begin{aligned} &\Delta f(H, G)_s - \frac{\Delta H_s}{G_{s-}} + \frac{f(H, G)_{s-}}{G_{s-}} \Delta G_s \\ &= \frac{H_s}{G_s} - \frac{H_{s-}}{G_{s-}} - \frac{\Delta H_s}{G_{s-}} + \frac{f(H, G)_{s-}}{G_{s-}} \Delta G_s \\ &= \frac{H_s}{G_{s-}(1 + \Delta A_s)} - \frac{H_{s-}}{G_{s-}} - \frac{\Delta H_s}{G_{s-}} + \frac{f(H, G)_{s-}}{G_{s-}} \Delta G_s \\ &= \frac{H_{s-}(\Delta A_s)(1 + \Delta A_s) - H_s \Delta A_s}{G_{s-}(1 + \Delta A_s)} \\ &= -\frac{\Delta N_{1s} \Delta A_s}{G_{s-}(1 + \Delta A_s)}, \end{aligned}$$

where A is a predictable process, and N is a local martingale. By the property of the Stieltjes integral, we have

$$\sum_{s \leq \cdot} \Delta f(H, G)_s - \frac{\Delta H_s}{G_{s-}} + \frac{f(H, G)_{s-}}{G_{s-}} \Delta G_s = -\frac{\Delta A}{G_-(1 + \Delta A)} \cdot N_1. \quad (2.8)$$

Thus $\left(e^{\lambda X}/\mathcal{E}(S(\lambda))\right)_t$ is a local martingale.

Since $e^x \geq x + 1$ and $e^{S(\lambda)_t} \geq \mathcal{E}(S(\lambda)_t)$,

$$\mathbb{E} \frac{e^{\lambda X_t}}{e^{S(\lambda)_t}} \leq \mathbb{E} \frac{e^{\lambda X_t}}{\mathcal{E}(S(\lambda)_t)} = 1. \quad (2.9)$$

Thus,

$$\begin{aligned} S(\lambda)_t &\leq \sum_{m=2}^{\infty} \frac{\lambda^m}{m!} C(W)_t^m \\ &\leq \frac{\lambda^2}{2(1-\lambda K)} C(W)_t. \end{aligned}$$

Set

$$A = \{X_t \geq x \text{ and } C(W)_t \leq y^2 \text{ for some } t\}, \quad (2.10)$$

$$\sigma = \inf\{t : X_t \geq x\}, \quad (2.11)$$

we have

$$\int_A \frac{e^{\lambda X_\sigma}}{e^{S(\lambda)_\sigma}} d\mathbb{P} \leq 1, \quad (2.12)$$

On A ,

$$\frac{e^{\lambda X_\sigma}}{e^{S(\lambda)_\sigma}} \geq \exp\left(\lambda x - \frac{\lambda^2 y^2}{2(1-\lambda K)}\right), \quad (2.13)$$

then

$$\mathbb{P}(A) \leq \exp\left(-\lambda x + \frac{\lambda^2 y^2}{2(1-\lambda K)}\right). \quad (2.14)$$

Take $\lambda = \frac{x}{y+Kx}$, we obtain

$$\mathbb{P}(X_t \geq x \text{ and } C(W)_t \leq y^2 \text{ for some } t) \leq \exp\left(-\frac{x^2}{2(y^2 + Kx)}\right).$$

Proof of Theorem 1.2

By Theorem 1.1, we can obtain

$$\mathbb{P}(|X^{\psi_1} - X^{\psi_2}| > u) \leq \exp\left(-\frac{u^2}{2(d_2^2(\psi_1, \psi_2) + d_1(\psi_1, \psi_2)u)}\right).$$

and

$$\mathbf{P}(|X^{\psi_1} - X^{\psi_2}| > ud_1(\psi_1, \psi_2) + \sqrt{u}d_2(\psi_1, \psi_2)) \leq \exp(-\frac{u}{2}).$$

for $u > 0$. We set $X^{\psi_0} = 0$.

Consider an admissible sequence (\mathcal{B}_n) such that

$$\sum_{n \geq 0} 2^n \Upsilon_1(B_n(\psi)) \leq 2\gamma_1(\Psi, d_1), \quad \forall \psi \in \Psi, \quad (2.15)$$

where $\Upsilon_1(B_n(\psi))$ is the diameter of the set $B_n(\psi)$ for d_1 , and an admissible sequence (\mathcal{C}_n) such that

$$\sum_{n \geq 0} 2^n \Upsilon_2(C_n(\psi)) \leq 2\gamma_2(\Psi, d_2), \quad \forall \psi \in \Psi, \quad (2.16)$$

where $\Upsilon_2(C_n(\psi))$ is the diameter of the set $C_n(\psi)$ for d_2 .

We may define partition \mathcal{A}_n for Ψ as follows: $\mathcal{A}_0 = \{\Psi\}$,

$$\mathcal{A}_n = \{B \cap C, B \in \mathcal{B}_{n-1}, C \in \mathcal{C}_{n-1}\}. \quad (2.17)$$

Consider a set Φ_n that contains exactly one point in \mathcal{A}_n . For $\psi \in \Psi$, $\pi_n(\psi)$ is the element of Φ_n that belong to $\mathcal{A}_n(\psi)$. We can easily obtain

$$X^\psi - X^{\psi_0} = \sum_{n \geq 1} (X^{\pi_n(\psi)} - X^{\pi_{n-1}(\psi)}). \quad (2.18)$$

Let Λ_n be the event defined by

$$|X^{\pi_n(\psi)} - X^{\pi_{n-1}(\psi)}| \leq u(2^n d_1(\pi_n(\psi), \pi_{n-1}(\psi)) + 2^{n/2} d_2(\pi_n(\psi), \pi_{n-1}(\psi))). \quad (2.19)$$

For $u > 1$, it easily follows

$$\mathbf{P}(\Lambda_n^c) \leq \exp(-u2^{n-1}).$$

Letting $\Omega_u = \cap_{n \geq 1} \Lambda_n$, we obtain

$$\mathbf{P}(\Omega_u^c) \leq p(u) := \sum_{n \geq 1} 2 \cdot 2^{2^{n+1}} \exp(-u2^{n-1}). \quad (2.20)$$

On Ω_u ,

$$|X^\psi - X^{\psi_0}| \leq u \sum_{n \geq 1} 2^n d_1(\pi_n(\psi), \pi_{n-1}(\psi)) + 2^{n/2} d_2(\pi_n(\psi), \pi_{n-1}(\psi)). \quad (2.21)$$

Hence,

$$\sup_{\psi \in \Psi} |X^\psi - X^{\psi_0}| \leq uS, \quad (2.22)$$

where

$$S := \sup_{\psi \in \Psi} \sum_{n \geq 1} 2^n (d_1(\pi_n(\psi), \pi_{n-1}(\psi)) + 2^{n/2} d_2(\pi_n(\psi), \pi_{n-1}(\psi))). \quad (2.23)$$

Thus

$$\mathbb{P}\left(\sup_{\psi \in \Psi} |X^\psi - X^{\psi_0}| > uS\right) \leq p(u) \quad (2.24)$$

Obviously,

$$p(u) \leq C \exp\left(-\frac{u}{2}\right). \quad (2.25)$$

When $n \geq 2$, we have $\pi_n(\psi), \pi_{n-1}(\psi) \in A_{n-1}(\psi) \subset B_{n-2}(\psi)$, so that

$$d_1(\pi_n(\psi), \pi_{n-1}(\psi)) \leq \Upsilon_1(B_{n-2}(\psi)). \quad (2.26)$$

Thus

$$\sum_{n \geq 1} 2^n d_1(\pi_n(\psi)) \leq C \sum_{n \geq 1} 2^n \Upsilon_1(B_{n-2}(\psi)) \leq C \gamma_1(\Psi, d_1). \quad (2.27)$$

Proceeding similarly for d_2 , we obtain

$$|X^\psi - X^{\psi_0}| \leq Cu(\gamma_1(\Psi, d_1) + \gamma_2(\Psi, d_2)) \quad (2.28)$$

in Ω_u . Thus

$$\mathbb{P}\left(\sup_{\psi \in \Psi} |X^\psi - X^{\psi_0}| > Cu(\gamma_1(\Psi, d_1) + \gamma_2(\Psi, d_2))\right) \leq C \exp\left(-\frac{u}{2}\right), \quad (2.29)$$

we complete the proof of (1.21).

We can obtain (1.22) through

$$\mathbb{E}|\xi| = \int_0^{+\infty} \mathbb{P}(|\xi| > u) du.$$

Remark 2.1. Let (Ψ, d) be a metric space, and let $(X^\psi, \psi \in \Psi)$ be a family of stochastic processes defined on a probability space (Ω, \mathcal{F}, P) . A primary problem is to study the bounds for $E \sup_{\psi \in \Psi} X^\psi$, where

$$E \sup_{\psi \in \Psi} S^\psi = \sup_{\psi \in \Psi} \left\{ E \sup_{\psi \in \Phi} S^\psi; \Phi \subset \Psi, \Phi \text{ finite} \right\}. \quad (2.30)$$

However, this is not easy at all for general processes. The generic chaining method was first invented by Talagrand in a series of articles to deal with $E \sup_{\psi \in \Psi} X^\psi$. In particular, under the increment condition

$$P(|X^{\psi_1} - X^{\psi_2}| > u) \leq 2 \exp \left(- \frac{u^2}{2d_2^2(\psi_1, \psi_2)} \right), \quad (2.31)$$

Talagrand [15] proved

$$P \left(\sup_{\psi \in \Psi} |X^\psi - X^{\psi_0}| > Cu\gamma_2(\Psi, d_2) \right) \leq C \exp \left(- \frac{u^2}{2} \right).$$

In addition, if the condition (2.31) is replaced by

$$P(|X^{\psi_1} - X^{\psi_2}| > u) \leq 2 \exp \left(- \min \left\{ \frac{u^2}{2d_2^2(\psi_1, \psi_2)}, \frac{u}{2d_1(\psi_1, \psi_2)} \right\} \right),$$

then it follows

$$P \left(\sup_{\psi \in \Psi} |X^\psi - X^{\psi_0}| > Cu(\gamma_1(\Psi, d_1) + \gamma_2(\Psi, d_2)) \right) \leq C \exp \left(- \frac{u}{2} \right). \quad (2.32)$$

Theorem 1.2 implies that if the following increment condition is satisfied

$$P(|X^{\psi_1} - X^{\psi_2}| > u) \leq \exp \left(- \frac{u^2}{2(d_2^2(\psi_1, \psi_2) + d_1(\psi_1, \psi_2)u)} \right),$$

then (1.21) still holds true.

3. Applications

In this section we shall first apply the previous results to functional index empirical processes. Consider a sequence of adapted stationary time series

$(Y_n)_{n \geq 0}$ on the discrete time stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbb{P})$. Let Ψ be the space of all bounded measurable functions in \mathbb{R} . For a $\psi \in \Psi$, define

$$X_n^\psi = \sum_{k=1}^n (\psi(Y_k) - \mathbb{E}(\psi(Y_k) | \mathcal{F}_{k-1})). \quad (3.1)$$

Obviously, for each ψ , X_n^ψ is a discrete time martingale. Note also X_n^ψ can be realized through a stochastic integral of ψ with respect to a multivariate point process. In fact, let $T_k = k$, $X_k = Y_k$, $\psi(k, x) = \psi(x_k)$, then

$$X_n^\psi = \psi * (\mu - \nu)_n. \quad (3.2)$$

A simple computation shows

$$\psi * \nu(dt, dx)_n = \sum_{k=1}^n \mathbb{E}[\psi(Y_k) | \mathcal{F}_{k-1}], \quad (3.3)$$

and

$$C(\psi)_n = \sum_{k=1}^n \mathbb{E}[(\psi(Y_k))^2 | \mathcal{F}_{k-1}]. \quad (3.4)$$

As a direct consequence of Theorem 1.1, we have

Theorem 3.1. *Suppose that, for all $k > 0$ and some $0 < K < \infty$*

$$\mathbb{E}[\max\{0, \psi(Y_k)\}^m | \mathcal{F}_{k-1}] \leq \frac{m!}{2} K^{m-2} \mathbb{E}[(\psi(Y_k))^2 | \mathcal{F}_{k-1}] \quad (3.5)$$

$$\mathbb{E}[\max\{0, \psi(Y_k)\} | \mathcal{F}_{k-1}] \leq K. \quad (3.6)$$

Then for each $x > 0$, $y > 0$,

$$P(|X_n^\psi| \geq x \text{ and } C(\psi)_n \leq y^2 \text{ for some } n) \leq \exp\left(-\frac{x^2}{2(xK + y^2)}\right). \quad (3.7)$$

Remark 3.2. *If we denote*

$$\Xi(\psi)_n = \mathbb{E}[\max\{0, \psi(Y_n)\} | \mathcal{F}_{n-1}], \quad (3.8)$$

and

$$Q(\psi, m)_n = \sum_{k=1}^n \mathbb{E}[\max\{0, \psi(Y_k)\}^m | \mathcal{F}_{k-1}], \quad m \geq 3. \quad (3.9)$$

The conditions (3.5) and (3.6) imply the conditions (1.14) in Theorem 1.1 for $\Xi(\psi)_n$ and $Q(\psi, m)_n$.

Furthermore, we define the metric for fix $n > 0$

$$d_1(\psi_1, \psi_2) = \max_{1 \leq k \leq n} \|\mathbb{E}[\max\{0, \psi_1(Y_k) - \psi_2(Y_k)\} | \mathcal{F}_{k-1}]\|_\infty, \quad (3.10)$$

$$d_2(\psi_1, \psi_2) = \max_{1 \leq k \leq n} \|\mathbb{E}[(\psi_1(Y_k) - \psi_2(Y_k))^2 | \mathcal{F}_{k-1}]\|_\infty^{1/2}. \quad (3.11)$$

where $\|\cdot\|_\infty$ stands for norm of L^∞ .

Theorem 3.3. *Suppose that, for all $\psi \in \Psi$ and some $0 < K < \infty$, (3.5) and (3.6) hold. Then*

$$\mathbb{P}\left(\sup_{\psi \in \Psi} |X_n^\psi| \geq Cu(\sqrt{n}\gamma_2(\Psi, d_2) + n\gamma_1(\Psi, d_1))\right) \leq C \exp(-\frac{u}{2}). \quad (3.12)$$

Proof. It follows directly from the proof of Theorem 1.2

$$\mathbb{P}\left(\sup_{\psi \in \Psi} |X_n^\psi| \geq Cu(\gamma_2(\Psi, \sqrt{n}d_2) + \gamma_1(\Psi, nd_1))\right) \leq C \exp(-\frac{u}{2}). \quad (3.13)$$

Note

$$\gamma_2(\Psi, \sqrt{n}d_2) = \sqrt{n}\gamma_2(\Psi, d_2), \quad \gamma_1(\Psi, nd_1) = n\gamma_1(\Psi, d_1) \quad (3.14)$$

Then (3.12) easily holds. \square

As a special example of functional index empirical processes, we consider the nonparametric maximum likelihood estimators below.

Let $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ be a family of probability measures, we assume that \mathcal{P} is dominated by a Lebesgue measure. Denote the density of P_θ by $f_\theta = \frac{dP_\theta}{dx}$, $\theta \in \Theta$. Fix a $\theta_0 \in \Theta$ such that $f_{\theta_0} > 0$, and let X_1, X_2, \dots be a sequence of i.i.d. observations from $P_0 = P_{\theta_0}$. Define the empirical distribution

$$P_n = \frac{1}{n} \sum_{k=1}^n \varepsilon_{(k, X_k)}(dt \times dx) \quad (3.15)$$

on the basis of the first n observations. The maximum likelihood estimator $\hat{\theta}_n$ of θ_0 is defined by

$$\int \log(f_{\hat{\theta}_n}) dP_n = \max_{\theta \in \Theta} \int \log(f_\theta) dP_n. \quad (3.16)$$

We assume throughout that a $\hat{\theta}_n$ exists.

It is very important to study the rate of convergence of $f_{\hat{\theta}_n}$ to f_θ in the theory of nonparametric statistical inference. Recall the Hellinger distance is usually used to describe the distance between two probability measures. In particular, for a pair (P, \bar{P}) of probability measures the Hellinger distance $H(P, \bar{P})$ is defined by

$$H^2(P, \bar{P}) = \frac{1}{2} \int \left(\sqrt{\frac{dP}{dQ}} - \sqrt{\frac{d\bar{P}}{dQ}} \right)^2 dQ \quad (3.17)$$

where Q is a measure dominating P and \bar{P} .

In our setting Q is a Lebesgue measure, $f = \frac{dP}{dx}$, $\bar{f} = \frac{d\bar{P}}{dx}$, and we simply write $h(f, \bar{f}) = H(P, \bar{P})$. It is natural to ask what the rate of convergence for $f_{\hat{\theta}_n}$ to f_{θ_0} in terms of $h^2(f_{\hat{\theta}_n}, f_{\theta_0})$. We have the following result in this aspect. Denote $\mathcal{G} = \{g_\theta := \sqrt{\frac{f_\theta}{f_{\theta_0}}} - 1, \theta \in \Theta\}$, and set

$$d_1(g_{\theta_1}, g_{\theta_2}) = E_{\theta_0} |g_{\theta_1}(X_1) - g_{\theta_2}(X_1)|, \quad (3.18)$$

$$d_2(g_{\theta_1}, g_{\theta_2}) = \sqrt{E_{\theta_0} (g_{\theta_1}(X_1) - g_{\theta_2}(X_1))^2}. \quad (3.19)$$

Theorem 3.4. *Suppose that there is a positive constant $0 < K < \infty$ such that for all $\theta \in \Theta$*

$$E_{\theta_0} |g_\theta(X_1)| \leq K, \quad E_{\theta_0} |g_\theta(X_1)|^m \leq \frac{m!}{2} K^{m-2} E_{\theta_0} |g_\theta(X_1)|^2, \quad m \geq 3. \quad (3.20)$$

Then it follows

$$P(h^2(f_{\hat{\theta}_n}, f_{\theta_0}) \geq u) \leq C \exp \left(- \frac{nu}{2C(\sqrt{n}\gamma_2(\mathcal{G}, d_2) + \gamma_1(\mathcal{G}, d_1))} \right). \quad (3.21)$$

Proof.

Since $\log(1+x) \leq x$ for any $x > -1$,

$$\begin{aligned}
0 &\leq \frac{1}{2} \int \log \frac{f_\theta}{f_{\theta_0}} dP_n - \frac{1}{2} \int (f_\theta - f_{\theta_0}) dx \\
&\leq \int \left(\sqrt{\frac{f_\theta}{f_{\theta_0}}} - 1 \right) dP_n - \frac{1}{2} \int (f_\theta - f_{\theta_0}) dx \\
&= \int \left(\sqrt{\frac{f_\theta}{f_{\theta_0}}} - 1 \right) d(P_n - P_0) + \int \left(\sqrt{\frac{f_\theta}{f_{\theta_0}}} - 1 \right) dP_0 - \frac{1}{2} \int (f_\theta - f_{\theta_0}) dx \\
&= \int \left(\sqrt{\frac{f_\theta}{f_{\theta_0}}} - 1 \right) d(P_n - P_0) + \int \sqrt{f_\theta f_{\theta_0}} dx - \frac{1}{2} \int (f_\theta + f_{\theta_0}) dx \\
&= \int \left(\sqrt{\frac{f_\theta}{f_{\theta_0}}} - 1 \right) d(P_n - P_0) - h^2(f_\theta, f_{\theta_0}).
\end{aligned}$$

Also, according to (3.16),

$$0 \leq \frac{1}{2} \int \log \left(\frac{f_{\hat{\theta}_n}}{f_{\theta_0}} \right) dP_n - \frac{1}{2} \int (f_{\hat{\theta}_n} - f_{\theta_0}) dx. \quad (3.22)$$

Thus we have

$$\begin{aligned}
&\mathbb{P} \left(h^2(f_{\hat{\theta}_n}, f_{\theta_0}) \geq Cu \left(\frac{1}{\sqrt{n}} \gamma_2(\mathcal{G}, d_2) + \frac{1}{n} \gamma_1(\mathcal{G}, d_1) \right) \right) \\
&\leq \mathbb{P} \left(\int \left(\sqrt{\frac{f_{\hat{\theta}_n}}{f_{\theta_0}}} - 1 \right) d(P_n - P_0) \geq Cu \left(\frac{1}{\sqrt{n}} \gamma_2(\mathcal{G}, d_2) + \frac{1}{n} \gamma_1(\mathcal{G}, d_1) \right) \right).
\end{aligned}$$

We now can complete the proof by Theorem 3.3. \square

Remark 3.5. *van de Geer [16, 17] discussed the similar problem on maximum likelihood estimators. To our knowledge, Theorem 3.4 is new in this area. We also remark that Theorem 3.4 can be extended by Theorem 1.2 and 3.3 to the stationary sample case with suitable maximum likelihood estimators. It is left to future work.*

Acknowledgments

This research work is support by the National Natural Science Foundation of China (No. 11371317, 11171303) and the Fundamental Research Fund of Shandong University (No. 2016GN019).

Reference

References

- [1] BENNETT, G. Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* , **57**, 33-45, (1962).
- [2] BENNETT, G. On the probability of large deviations from the expectation for sums of bounded independent random variables. *Biometrika* , **50**, 528-535, (1963).
- [3] BERCU, B., DELYON, B. AND RIO, E. *Concentration inequalities for sums and martingales. SpringerBriefs in Mathematics*, Springer, (2015)
- [4] BENTKUS, V. An inequality for tail probabilities of martingales with differences bounded from one side. *J. Theoret. Probab.* , **16**, 161-173, (2003).
- [5] BENTKUS, V. On Hoeffding's inequalities. *Ann. Probab.* , **32**, 1650-1673, (2004).
- [6] BERNSTEIN, S.N. *Theory of Probability*, Moscow. (1927) .
- [7] FAN, X., GRAMA, I. AND LIU, Q. Exponential inequalities for martingales with applications. *Electron. J. Probab.* , **20**, (2015).
- [8] FREEDMAN, D. On tail probabilities for martingales. *Ann. Probability* , **3**, 100-118, (1975).
- [9] Hoeffding, W Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* , **58**, 13-30, (1963).
- [10] JACOD, J. AND SHIRYAEV, A. *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* **288**. Springer, Berlin, (2003).

- [11] LEDOUX, M. *The concentration of measure phenomenon, Mathematical Surveys and Monographs* 89. American Mathematical Society, Providence, RI, (2001).
- [12] LEDOUX, M. AND TALAGRAND, M. *Probability in Banach spaces. Isoperimetry and processes.*, Springer, Berlin, (1991).
- [13] SHORACK, G. WELLNER, J. *Empirical Processes with Applications to Statistics*, Wiley, New York. (1986).
- [14] TALAGRAND, M. New concentration inequalities in product spaces. *Invent. Math.*, **126**, 505-563, (1996).
- [15] TALAGRAND, M. *Upper and Lower Bounds for Stochastic Processes: Modern Method and Classical Problems*, Springer, Berlin, (2014).
- [16] VAN DE GEER, S. Hellinger-consistency of nonparametric maximum likelihood estimators. *Ann. Statist.*, **21**, 14-44, (1993).
- [17] VAN DE GEER, S. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. *Ann. Statist.*, **23**, 1779-1801, (1995).